02: Sampling

Stat 120 | Fall 2025

[List your group here]

Press the "play" button below to run the chunk of code. This (1) loads the libraries that we need and (2) tells R to read `mission_data.csv` from my website folder into your R session and call it `mission_data`.

```
library(tidyverse)
#mission_data = read_csv("https://www.math.carleton.edu/aluby/stat120/mission_data.csv")
mission_data = read_csv("../data/mission_data.csv") # update to server data!!
```

Check your "environment" pane in the upper right to make sure you can see a dataset called `mission_data`. Try clicking it, or running `View(mission_data)` in the console to bring up the data viewer.

spoiler alert: The next chunk of code computes the *population mean*.

```
mean(mission_data$length)
```

```
[1] 5.947802
```

```
set.seed(091824) # Sets the random seed so we all get the same answer
sample = sample(1:nrow(mission_data), size = 10) # selects a random sample of size 10 from the r
sample # prints the sample
```

```
 [1] 213 328 237 164  34 311 115 121 253 250
```

The next chunk of code `slices` our population to draw our sample. Note that the `position` variable should match the `sample` output above.

```
mission_sample = mission_data %>%
  slice(sample)

mission_sample
```

```
# A tibble: 10 x 4
   paragraph word     position length
       <dbl> <chr>       <dbl>  <dbl>
 1         3 be            213      2
 2         4 of            328      2
 3         4 the           237      3
 4         3 carleton      164      8
 5         1 learning       34      8
 6         4 for           311      3
 7         3 the           115      3
 8         3 students      121      8
 9         4 arts          253      4
10         4 in            250      2
```

```
mean(mission_sample$length)
```

```
[1] 4.3
```

To try other random samples, change (or remove!) the `set.seed()` line of code, and try re-running the rest of the code. Do you ever get a sample mean that looks like your "by hand" sample mean?

When you're done, **knit this file** and try uploading the PDF to gradescope. There are two questions, one for the *population mean* and one for a *sample mean*. Be sure to mark the pages so I can see your answers!